Graspability を導入した DCNN による物体把持位置検出

○荒木諒介　長谷川昂宏　山内悠嗣　山下隆義　藤吉弘亘 (中部大学)
堂前幸康　川西亮輔　関真規人 (三菱電機 (株))

1.　はじめに

産業用ロボットや生活支援ロボットの重要なタスクとして，工業部品や日用品などの物体を正確に把持する (掴む) ことが挙げられる．このピッキングタスクを実現するには，ロボットに搭載されているカメラセンサを用いて物体の画像を撮影し，その物体の最適な把持位置を検出する必要がある．ピッキングロボット向けの把持位置検出法は，これまでに多くの手法が提案されており [1]-[6]，機械学習を用いる手法と用いない手法に大別できる．近年では，特に Deep Learning を用いた把持位置検出法が注目されている．

本研究では，把持可能性 (Graspability) を導入した Deep Convolutional Neural Network(DCNN) による物体の把持位置検出法を提案する．提案手法の主な特徴として以下が挙げられる．

1. Deep Convolutional Neural Network を利用した把持位置の検出
2. Heterogeneous Learning を利用した複数タスクの学習
3. 5 次元表現を利用した正確な把持位置の取り扱い
4. Graspability を導入した高精度な把持位置検出

2.　関連研究

本章では，把持位置検出の従来法について述べる．

2.1　2 段階 DNN による把持位置検出法

Lenz 等により提案された 2 段階の Deep Neural Network (DNN) を用いた把持位置検出法 [1] は，規模の異なる 2 つの DNN を用意して二段階の処理により把持位置を検出する．図 1 に示すように，まずは 1 段階目で Small Neural Network を用いて複数の把持位置の候補点を検出する．2 段階目では 1 段階目より多くのユニットを使用した Large Neural Network を用いて，複数の把持位置候補を 1 つに絞り込む．1 段階目を小規模なネットワークで，2 段階目を大規模なネットワークで全体のネットワークを構築することにより，計算コストを抑えている．しかし，アルゴリズムが複雑化し，様々な矩形の方向やサイズを考慮した複数回のラスタスキャンを必要とするという欠点がある．

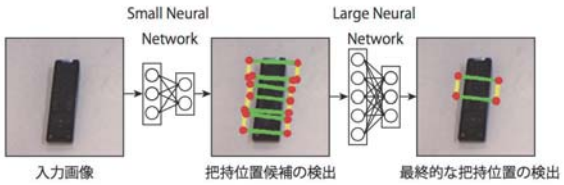


図 1　2 段階 DNN を利用した把持位置検出

2.2　ハンドモデルのテンプレートを利用した把持位置検出

機械学習を利用しない把持位置検出法 [3]-[6] の 1 つとして，Domae 等により提案された Fast Graspability Evaluation (FGE) [3] が挙げられる．FGE は，ロボットのハンドモデルと物体領域の 2 値画像を畳み込むことで把持位置を検出する．ロボットハンドが物体に接触する領域と衝突する領域の 2 値パターンをそれぞれテンプレートとして保持する．これらのテンプレートを対象の物体の 2 値画像と畳み込んだ後，ガウシアンフィルタをかけて Graspability マップを生成する．様々なハンドモデル，回転角，開き幅のテンプレートを用いて以上の処理を行うことにより，最適な把持位置とハンドの状態を検出する．FGE の処理の流れを図 2 に示す．FGE では，ハンドの状態毎に画像の畳み込みを行う必要があるため，計算コストが増加するという欠点がある．

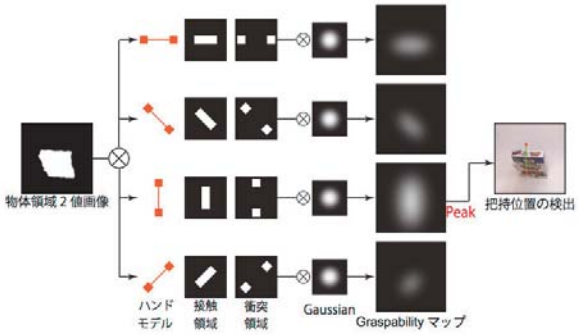


図 2　Fast Graspability Evaluation による把持位置検出

3.　提案手法

本研究では，Deep Convolutional Neural Network (DCNN) と Heterogeneous Learning を利用して，物体把持位置と把持可能性 (Graspability) を同時に検出する．把持位置と Graspability を同時に求めることで，高精度な物体把持位置検出が可能となる．本章では，提案手法の詳細について述べる．

3.1　Deep Convolutional Neural Network

Deep Convolutional Neural Network (DCNN) は，畳み込み層とプーリング層を複数層繰り返して構成された Neural Network である．畳み込みとプーリングを繰り返し得られた特徴マップを全結合層へ入力して，回帰や識別結果が得られる．

畳み込み層では，入力された画像及び特徴マップにフィルタサイズ $n \times n$ の重みフィルタを畳み込み，得

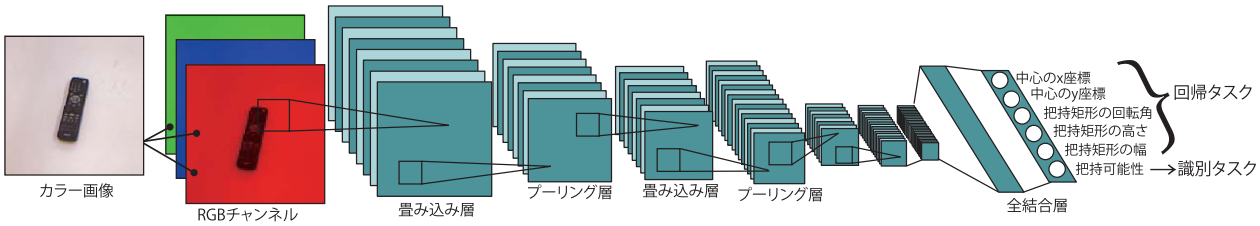


図 3　提案手法の DCNN の構成

られた値 v を活性化関数 $f(v)$ へ入力する．提案手法では，活性化関数に Rectified Linear Unit (ReLU) を用いる．ReLU は式 (1) のように表される関数であり，v が負の値ならば 0 を返し，正の値ならば v の値をそのまま返す．シグモイド関数などで起こる，v の値が大きい時に勾配が消える問題は発生しない．

$$f(v) = \max(0, v) \tag{1}$$

その後，ReLU によって得られた結果を特徴マップとして格納する．

プーリング層では，畳み込み層で得られた特徴マップを縮小する処理として Max Pooling を用いる．

提案手法で用いる DCNN の構成の外観図を図 3 に，詳細な構成を表 1 に示す．

表 1　提案手法で使用する DCNN の詳細

Layer	詳細
畳み込み層 1 層目	畳み込みフィルタ：9 × 9 活性化関数：ReLU プーリング：2 × 2
畳み込み層 2 層目	畳み込みフィルタ：9 × 9 活性化関数：ReLU プーリング：2 × 2
畳み込み層 3 層目	畳み込みフィルタ：7 × 7 活性化関数：ReLU プーリング：2 × 2
畳み込み層 4 層目	畳み込みフィルタ：5 × 5 活性化関数：ReLU プーリング：なし
全結合層 1 層目	ユニット数：2000
全結合層 2 層目	ユニット数：6

3.2　Heterogeneous Learning

提案手法では，Heterogeneous Learning により回帰タスクである「把持位置の座標」と識別タスクである「Graspability」を単一の DCNN で学習する．これにより，検出の際には把持位置の座標と併せて，その把持位置が「把持にふさわしいかどうか (把持可能性) 」を得ることができる．

3.2.1　把持位置の 5 次元表現

把持位置は，4 つの頂点からなる矩形 (長方形) で表現する．しかし，把持位置を示す矩形の 4 つの頂点の座標をそれぞれ学習すると，検出時に把持位置を必ずしも長方形として検出できないという問題がある．そこで，本研究では Lenz ら [1] が用いた Five-Dimensional Representation (5 次元表現) により把持位置を取り扱う．この表現方法は，物体把持位置の矩形を中心の x 座標，中心の y 座標，把持矩形の回転角 (θ)，把持矩形の高さ (height)，把持矩形の幅 (width) の 5 つの要素から成る．5 次元表現での把持位置の表示例を図 4 に示す．角度の範囲は $0 \sim \pi$(rad) であり，それを超える場

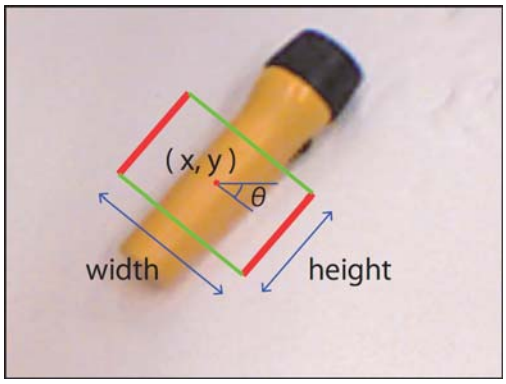


図 4　5 次元表現を利用した物体把持位置

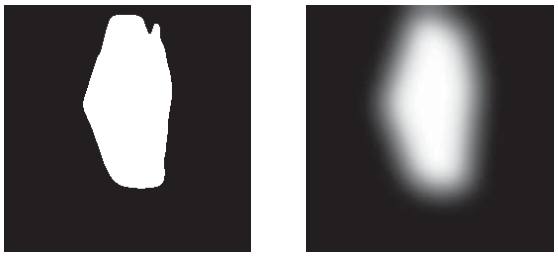
合は π の剰余の値となる．また，図中赤色の部分がロボットハンドの位置となるため，高さはロボットハンドの長さ，幅はロボットハンドの開き幅となる．5 つの要素すべてにおいて負の値は取り得ないため，DCNN の活性化関数として ReLU を用いることが可能である．

3.2.2　Graspability の導入

本研究では，把持位置の検出と同時に，その把持可能性 (Graspability) を求める．Graspability は，本当に把持にふさわしいかどうか，把持できるかどうかを表すもので，0 ～ 1 の値で表現する．「物体ではない部分」「物体であるが，把持できない把持位置座標」を 0，「物体であり，把持が可能である把持位置座標」を 1 として学習する．

しかし，把持位置が物体上にあり把持可能であっても，その把持位置が物体の端の場合は，把持した後に落下する可能性も考えられる．そこで，このような把持位置を避けるべく，物体領域の中心の Graspability を高くする．提案手法では，図 5(a) のように物体領域を二値画像で表現して，その二値画像に対して図 5(b) のようにガウシアンフィルタをかけて作成した Graspability マップを教師信号に用いる．これにより，物体の中心に近い把持位置は 1.0 となり，中心から遠ざかるほど

0.5 に近い値となる．ただし，把持できない把持位置は物体の中心に近い場合でも 0 とする．



(a) 物体の二値画像　(b) Graspability マップ

図 5　物体の二値画像と Graspability マップ

図 6 に Graspability 値の例を示す．図 6(a) では，把持できない把持位置を示しているので，Graspability は 0.0 になる．図 6(b) では，物体の端の把持位置を示しているので，Graspability は 0.5 になる．図 6(c) では，物体の中心に近い把持位置を示しているので，Graspability は 1.0 になる．

(a) 0.0　(b) 0.5　(c) 1.0

図 6　把持位置による Graspability 値の違い

3.3　把持位置の検出

入力画像から把持位置を検出するには，250×250 画素の検出ウィンドウをラスタスキャンする．検出ウィンドウごとに，把持位置の候補が検出される．提案手法では，把持位置を検出するとき把持矩形の回転角度や高さ，幅をを回帰により求めるため，回転角度やスケールを変化したラスタスキャンを必要としない．これらの把持位置候補から最終的な把持位置を決定するために，先述した Graspability を用いる．本研究では，単純に Graspability が最も高い把持位置を採用する．

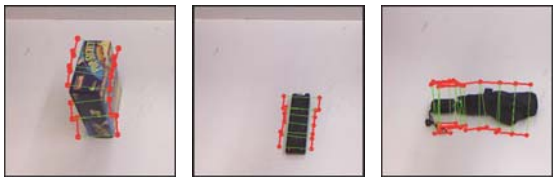
4.　評価実験

本研究で提案する手法の有効性を確かめるために，評価実験を行う．比較手法は 2 段階 DNN による把持位置検出法 [1] を用いる．

4.1　データセット

本研究では評価実験用のデータセットとして，Cornell University の Robot Learning Lab が公開している Cornell Grasping Dataset [1] を使用する．このデータセットは，日用品，食材，OA 機器，調理器具など 280 種類のアイテムの RGB 画像が 885 枚存在する．また，各 RGB 画像に対して，距離点群データ，Positive 教師信号，Negative 教師信号が用意されている．教師信号は Positive，Negative ともに 1 枚につき複数存在する．このデータセットの画像 644 枚を学習し，学習に用いていない画像 224 枚を評価実験に用いる．

データセットに含まれる物体の RGB 画像とその Positive 教師信号と Negative 教師信号の例を図 7 に示す．



(a) Positive 教師信号

(b) Negative 教師信号

図 7　Cornell Grasping Dataset

4.2　評価方法

把持位置の検出成功の判定は，正解把持位置と検出結果の矩形の重なり率を用いる．式 (2) の条件を満たした場合に検出成功，それ以外を検出失敗とする．

$$\frac{R_d \cap R_t}{R_d \cup R_t} \geq T \tag{2}$$

ここで，R_d は検出した把持位置の矩形領域，R_t は Positive の教師信号の把持位置の矩形領域である．なお，教師信号は 1 枚の画像につき複数あるため，Positive の教師信号すべてにおいて式 (2) で判定して，1 つでも条件を満たす場合，検出成功とする．

4.3　実験結果

式 (2) の閾値 T の値を 0.2 から 0.5 刻みで 0.45 まで変化させた時の検出率 [%] を図 9 に示す．

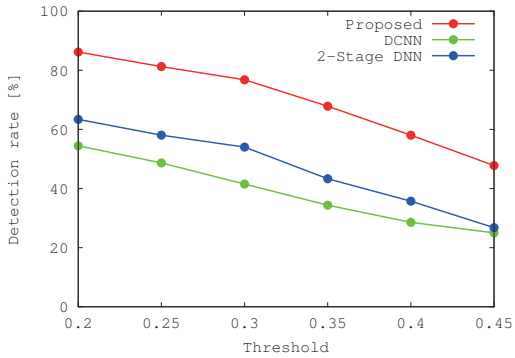


図 9　各手法による把持位置検出の結果とその比較

青色は 2 段階 DNN[1] による検出結果，緑色は DCNN による検出結果，赤色は提案手法による検出結果を示す．DCNN は提案手法と同じ構成のネットワークであり，畳み込み層は 4 層，全結合層は 2 層である．提案手法は DCNN に Graspability を導入し，Heterogeneous

[1] http://pr.cs.cornell.edu/grasping/rect_data/data.php

図 8　提案手法と従来手法による把持位置検出結果と正解把持位置

Learning で学習したものである．図 9 より，提案手法は従来手法の 2 段階 DNN と比べ 22.8 %検出率を向上させることができた．また，DCNN と比べ，Graspability を導入することで 30.9 %検出率を向上させることができた．ラスタスキャンによる検出ウィンドウ数が 240 個のとき，検出時間は 1 枚あたり約 30 秒である．

提案手法と 2 段階 DNN による把持位置検出例を図 8 に示す．赤色と緑色の線分で囲まれた矩形が検出結果および正解把持位置である．赤色の線分はロボットハンドの位置にあたる．提案手法の把持位置は物体領域の中心付近を検出しており，安定した把持が期待できる．

5.　おわりに

本研究では，DCNN による Graspability を導入した高精度な物体把持位置検出法を提案した．DCNN を利用して把持位置の候補を検出するだけではなく，Graspability を導入することで適切な把持位置検出を実現した．今後は，実際のロボットを用いた強化学習による物体把持位置の高精度化，把持に最適なロボットハンドの種類の検出，PointCloud 等を用いた三次元ベースでの把持位置検出に取り組んでいく．

謝辞

この成果は，国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです．

参　考　文　献

[1] I. Lenz, H. Lee, and A. Saxena, “Deep Learning for Detecting Robotic Grasps”, International Journal of Robotics Research, vol.34, no.4-5, pp.705-724, 2015.

[2] J.Redmon, and A.Angelova,“Real-Time Grasp Detection Using Convolutional Neural Networks”, International Conference on Robotics and Automation, pp.1316-1322, 2015.

[3] Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, and T. Hirai, “Fast graspability evaluation on single depth maps for bin picking with general grippers”, International Conference on Robotics and Automation, pp.1997-2004, 2014.

[4] K. Harada, K. Nagata, T. Tsuji, N. Yamanobe, A. Nakamura, and Y. Kawai, “Probabilistic approach for object bin picking approximated by cylinders”, International Conference on Robotics and Automation, pp.3742-3747, 2013.

[5] M. Nieuwenhuisen, D. Droeschel, D. Holz, J. Stuckler, A. Berner, J. Li, R. Klein, and S. Behnke, “Mobile bin picking with an anthropomorphic service robot”, International Conference on Robotics and Automation, pp.2327-2334, 2013.

[6] B. Drost, M. Ulrich, N. Navab, and S. Ilic, “Model globally, match locally: Efficient and robust 3D object recognition”, Computer Vision and Pattern Recognition, pp.998-1005, 2010.

[7] 長谷川昂宏, Xuanyi Sheing, 荒木諒介, 山内悠嗣, 山下隆義, 藤吉弘亘, “Heterogeneous Learning によるオブジェクトネスと物体把持位置の検出”, 画像センシングシンポジウム, 2016.